

Evaluation of Pointing Navigation Interface for Mobile Robot with Spherical Vision System

Kyohei YOSHIDA, Fuminori HIBINO, Yasutake TAKAHASHI, and Yoichiro MAEDA
Dept. of Human and Artificial Intelligent Systems, Graduate School of Engineering, University of Fukui
3-9-1, Bunkyo, Fukui, Fukui, 910-8507, Japan
Email: {kyosida,fhibino,yasutake,maeda}@ir.his.u-fukui.ac.jp

Abstract—In human robot interaction, intuitive interface is necessary. A specific interaction device, for instance, a joystick or a teaching pendant, is not usually intuitive and needs trainings for a general user. Instruction by gesture is one of the intuitive interfaces and a potential user does not need any training for showing a gesture. Pointing is one of the simplest gestures. Hibino et. al.[1] proposed a simple human pointing recognition system for a mobile robot that has an upward directed camera and recognizes human pointing and navigate itself to the place a user is pointing by simple visual feedback control. This paper shows improvement of the method and investigates the validity and usefulness of the proposed method with questionnaire investigations with the proposed and conventional user interfaces.

Keywords—pointing navigation interface; spherical vision system; omni-directional mobile robot

I. INTRODUCTION

Recently, robots are getting popular not only in factory environments but also in human daily life environments. Especially, entertainment robots, for instance, AIBO from Sony[2] and Nao from Aldebaran[3], attract the general people. On the other hand, more practical mobile robots becomes gradually popular, for example, rumba from iRobot[4] and SO-ZI from CCP[5], as a vacuum cleaner robot. More practical robots interacting with a general user are going to be popular in near future although the current vacuum cleaner robots work automatically without human interaction.

In human robot interaction, intuitive interface is necessary. A specific interaction device, for instance, a joystick or a teaching pendant, is not usually intuitive and needs trainings for a general user. Instruction by gesture is one of the intuitive interfaces and a potential user does not need any training for showing a gesture.

Most conventional studies related to gesture recognition assume that height of camera position is about the one of human torso[6], [7], [8]. This assumption is reasonable for a robot that is as tall as a human, however, it is not for a smaller robot like a vacuum cleaner robot for a general user, for example, rumba.

On the other hand, localization and navigation with an upward directed camera on a mobile robot has been studied[9], [10], so far. The mobile robot watches ceiling above and localizes by itself. The method has some advantages over the other vision-based localization methods. One of them is that ceiling image can be captured with less occlusion than side view. Another one is less computational cost for localization based on image processing. A ceiling image has almost only

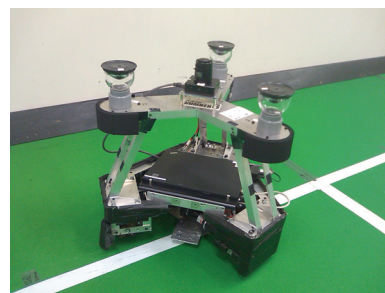


Fig. 1. Mobile Robot with a Spherical Vision System consists of an Upward Directed Fish-Eye Camera and Cameras with Omni-Mirrors

parallel and rotation transform with less magnification transform while the mobile robot moves on the flat. Therefore, visual processing can eliminate an image magnification factor. For instance, template matching needs less computational cost if it does not have to deal with the image magnification factor.

A fish-eye upward directed camera has not only same advantages with the conventional upward-directed-camera based localization methods but also wide view angle. The camera captures not only ceiling but also a human as a potential general user, simultaneously. Therefore, a robot with a fish-eye upward directed camera can localize itself with ceiling image while it recognizes a human gesture.

Hibino et. al.[1] proposed a simple human pointing recognition system for a mobile robot. The mobile robot with a fish-eye upward directed camera recognizes human pointing and navigates itself to the place the human points with a simple visual feedback controller. A human user can navigate the mobile robot intuitively by pointing at desired position on a field with his hand. However, the human pointing system proposed by Hibino et. al.[1] has low operability where the robot goes far from the user. This paper shows improvement of the proposed method with a simple visual feedback control system. Real robot experiments and questionnaire investigation with 11 persons in their twenties are conducted to show the validity and usefulness of the proposed method.

II. MOBILE ROBOT WITH SPHERICAL VISION SYSTEM

Figure 1 shows an omni-directional vehicle with a spherical vision system consists of an upward-directed fish-eye camera and cameras with omni-mirrors, which we designed and built. The upward-directed fish-eye camera captures a sight upper

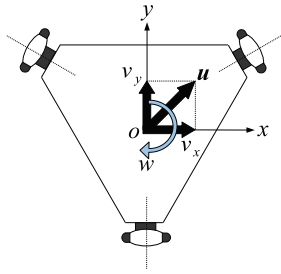


Fig. 2. Robot Coordinate

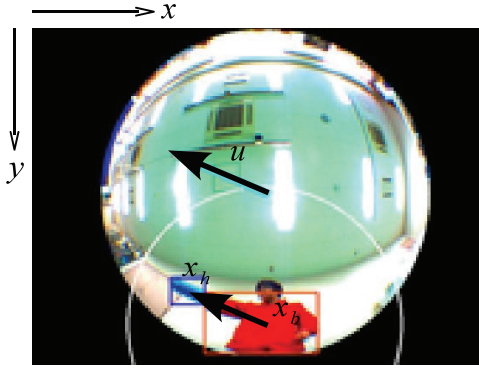


Fig. 3. Simple Visual Feedback for Navigation

than horizon while cameras with omni-mirrors capture sight lower than horizon. Height of the robot is about 50cm and the cameras capture a human surrounding the robot. The method proposed in this paper focuses on the captured images by the upward-directed fish-eye camera. Figure 7 shows captured images. For simplicity of a vision system, a human wears a red jacket and a blue glove.

Figure 2 shows a model of our omni-directional vehicle with three omni-wheels. Coordinates x and y indicate left and front directions of the robot. The robot has two omni-wheels at left and right front and the other at back. It can move toward any direction v_x, v_y including rotation w on a plane surface.

III. RECOGNITION OF HUMAN POINTING AND NAVIGATION

Recognition of human pointing is realized with a simple algorithm. A vision system extracts red and blue regions on the captured image to recognize positions of the human torso and hand. In order to reduce hand recognition error, blue regions within a certain distance from a red/torso region (shown by white circles in Figure 7) are regarded as candidates of hand region. The largest region of them is recognized as a hand.

Navigation algorithm is designed as a simple visual feedback. The robot moves to the position so that the hand comes to the human torso region on the fish-eye camera image. The human hand comes to the torso area where the robot locates at the place human is pointing as shown in Figure 7(e). The hand position is far from the torso region when the human points with his hand at a position where the robot does not locate as shown in Figure 7(b). The intuitive idea leads to a simple visual

feedback based navigation controller. The movement vector $u = (v_x, v_y)$ for the navigation is designed by Hibino et al.[1] as followings:

$$u = k(x_h - x_b) \quad (1)$$

while k , x_h , and x_b are a control gain and center position vectors of the extracted human hand and torso on the captured image, respectively (see Figure 3).¹ Actually an object located at right side is projected to left side on a ceiling image and vice versa, therefore, x direction on a ceiling image is reversed when it calculate the movement vector u . Furthermore, in general image processing, y coordinate is set from top to down as shown in Figure 3. We just flip y coordinate upside down in order to follow the robot coordinate shown in Figure 2. This means that upper and right directions on the camera image correspond to forward and left in the robot coordinate.

However, the controller designed by Eq.1 has low operability where the robot goes far from the user. The region of the user in the image captured by the upward-directed fish-eye camera becomes smaller if the user is farther from the robot. This means Eq.1 gives smaller movement vector when the user is farther from the robot while it is larger vector when the user is nearer to the robot. In order to compensate the length of the movement vector depending on the distance from the robot and the user, the movement vector is magnified based on the scale of the user region in the image. If the user is projected small in the image, the image is virtually magnified so that the user region in the image becomes same size of the user region projected in the image when the user is near to the robot. This means that the user size in the image is virtually kept as constant so that the length from the torso and the hand of the user keeps constant if the user shows exact same pointing gesture to the robot. The concrete adaptive gain control is designed as followings:

$$u = k \frac{f(S_{\max})}{f(S)} (x_h - x_b) \quad (2)$$

where S , $S_{\max}$, $f(S)$ are red region size in the current image, the red region size when the user stands nearest to the robot, and length of the vector between the red(torso) and blue(hand) regions, respectively. $\frac{f(S_{\max})}{f(S)}$ magnifies the moving vector as long as the user stands nearest to the robot. Data of relationship between the red region size and the length of vector between red and blue regions while the user is pointing horizontally as shown at bottom right in Figure 4 are collected beforehand. The data is approximated with the function Eq.3.

$$f(S) = aS^2 + bS + c \quad (3)$$

Parameters, a , b , and c , of the function $f(S)$ are tuned based on the data. Figure 4 shows the data and the function with tuned parameters.²

¹ Rotation speed w is fixed to zero because it is not directly related to the navigation in experiments.

² $a = -0.000000234337$, $b = 0.0111673$, $c = 39.9448$

IV. PERFORMANCE EVALUATION OF HUMAN POINTING NAVIGATION

First of all, control performance of the proposed human pointing navigation is evaluated. Experiments are carried out on a small soccer field in order to evaluate the performance of the proposed method. Figure 5 shows a experimental setup for evaluation of human pointing and navigation. The yellow and blue regions indicate goals. A human stands on (0, 0) depicted in Figure 5 and faces the opposite side line. The human points with his right hand at one point on one of the dots depicted in Figure 5. An omni-directional mobile robot with an upward-directed fish-eye camera recognizes the human pointing and navigates itself to the pointed place. Errors between the pointed place and the robot destination are recorded.

Figure 6 shows how the robot works with human pointing. While the human does not show his hand to the robot, the robot stays around the human (Figure 6(a)). When the human points with his right hand at a certain point on the field, the robot recognizes hand and torso region on the fish-eye camera image and moves to the point successfully (Figures 6(b)(c)(d)(e)).

In order to evaluate navigation performance, error between a query point and converged location of the navigated robot is investigated. A human points at one of the dots depicted in Figure 5. The size between the dots is 0.5 [m] by 0.5 [m]. Figure 8 show the query points on the field and the final locations of the navigated robot based on the controller with adaptive gain control. Red crosses and heads of green arrows indicate the query points and the final locations of the navigated robot, respectively.

The error becomes small when the query point is close to the pointing user because pointing vector is detected large while the robot is close to the human. The controller with the adaptive gain control (Figure 8) shows less errors when the robot starts from the region [-2m to 2m] horizontally. Unfortunately, the error becomes over 0.5[m] if the pointed place is over 2m from the user. The reason is that the red (torso

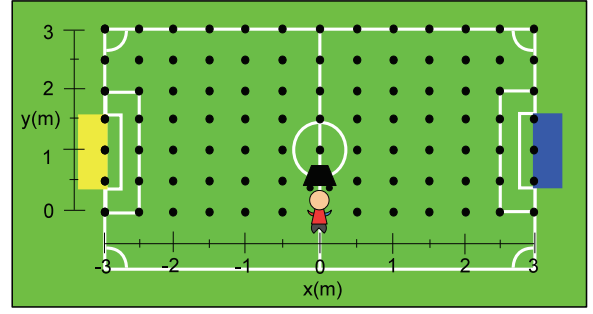


Fig. 5. Experimental Setup for Evaluation of Human Pointing Navigation

of the user) region in the image becomes too small when the robot goes far from the user to compensate the moving vector accordingly.

The errors on the right side on the field are smaller than the ones on the left side. One of the reasons is that the projection of the human right hand to the fish-eye camera image tends to be smaller while the human points left side of the field. Another reason is that the center of the torso is not proper position of origin of the moving vector as described in Eqs. 1 and 2. Actually, the position of the right shoulder should be the position of the origin when the user points with his right hand. When the user points in right side, the center of the torso is close to the position of the right shoulder but it becomes far if the user points in the left side.

Figure 9 shows a comparison on the error rate between the adaptive gain control and the fixed one[1]. A user stands at [0m, 0m] on the field and point [0m, 1m] with his right hand. The robot starts to navigate to the point from one of points depicted with red crosses. It is apparent that the adaptive gain control leads the robot to more accurate point than the fixed one. The fixed gain control shows larger errors vertically because the size of vector on the projected image by the fish-eye camera, that is, $x_h - x_b$ in Eqs. (1) and (2), tends to be too small to lead the robot to the query point. On the other hand, the adaptive gain control compensates the effect by tuning the gain accordingly.

V. USABILITY EVALUATION OF HUMAN POINTING NAVIGATION

In order to evaluate usability of the proposed human pointing navigation, questionnaire investigation with 11 persons in their twenties are conducted. We prepare 4 methods for robot navigation: Pointing with fixed control gain, pointing with adaptive gain control (Figure 10(a)), a joystick control (Figure 10(b)), and a game-pad one (Figure 10(c)). A subject is asked to navigate the robot as he/she like with one of the method for about half a minute to become used to the method. After he/she becomes used to the method, he/she navigates the robot along a 2.1m by 1.5m rectangle on the soccer field for about one minute. He/she repeats the procedure for all methods. Then, he/she answers questionnaires as shown below:

- Amusement: no fun(1) - fun(5)
- Operability: hard(1) - easy(5)

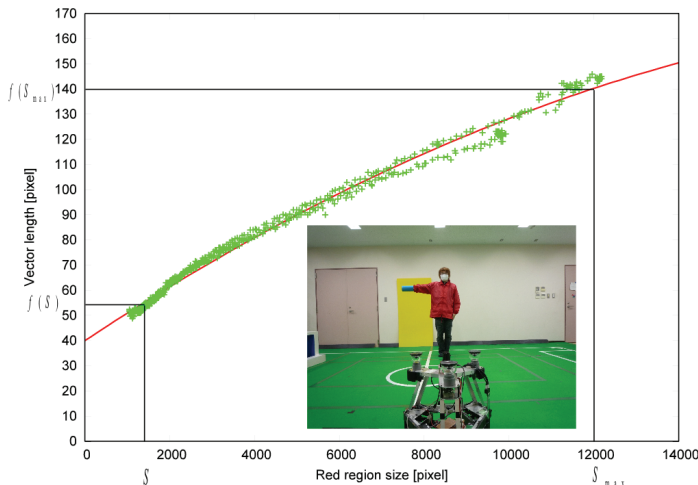
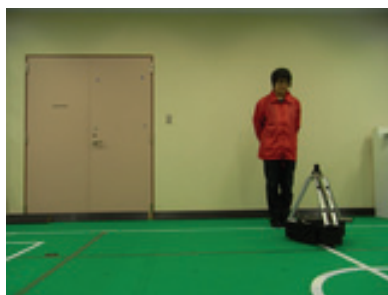
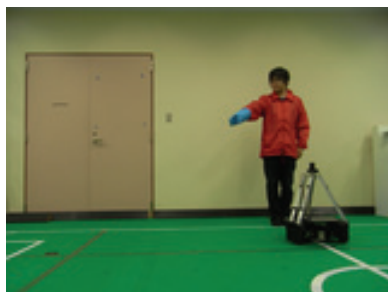


Fig. 4. Relationship between Red Region Size and Length of Vector between Red and Blue Regions



(a)



(b)



(c)

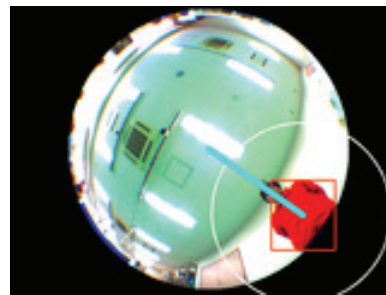


(d)

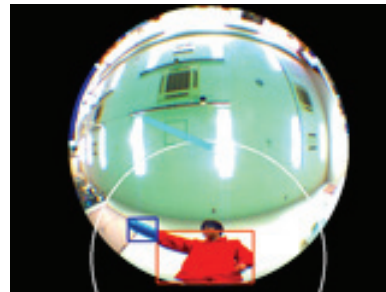


(e)

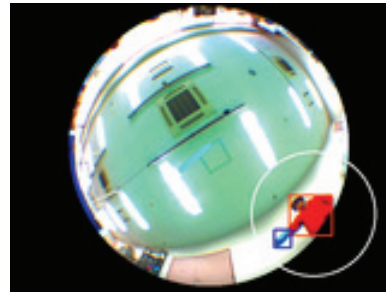
Fig. 6. Human Pointing and Navigation



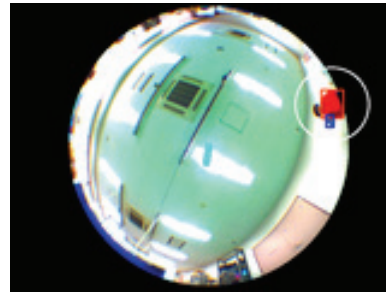
(a)



(b)



(c)



(d)



(e)

Fig. 7. Images from the Upward-Directed Fish-Eye Camera

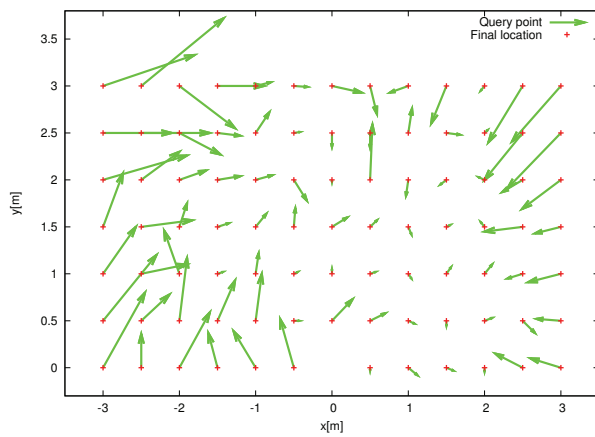


Fig. 8. Navigation Errors with Adaptive Gain Control

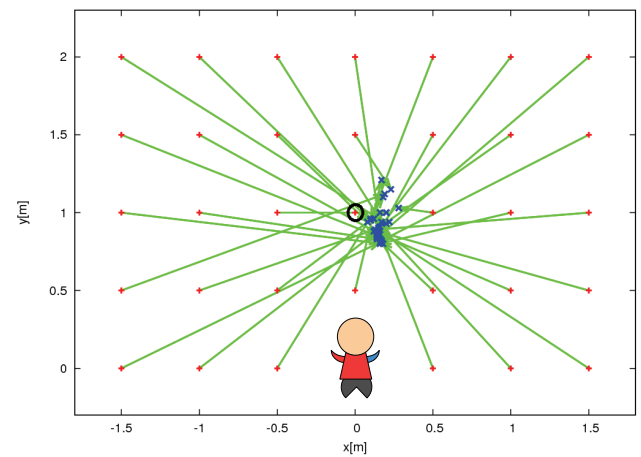
- Favorite: dislike(1) - like(5)

Figures 11, 12, and 13 show the questionnaire results. Vertical axes indicate number of votes for each degree (from 1 to 5) of the items (amusement, operability, and favorite). The most subjects are used to the gamepad because they sometimes play computer games but not used to the joystick. They feel fun with the joystick and the pointing with fixed/adaptive control gain because of new experiences for them. They feel hard to navigate the robot with the joystick and the pointing with fixed control gain because they are not used to the joystick and the the pointing with fixed control gain makes the robot behave in different way depending on the distance from the robot to the user. On the other hand, the gamepad shows the best operability because they are used to it and the pointing with adaptive control gain shows the 2nd best. Eventually, the pointing with adaptive control gain shows the best favorite evaluation over all. The reason is that the subjects enjoy the new experience and the operability is well nevertheless they are not used to the pointing navigation interface.

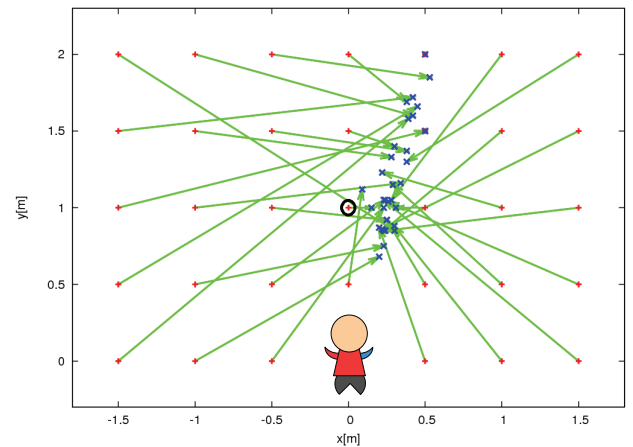
VI. CONCLUSIONS AND FUTURE WORK

We proposed a simple human pointing recognition system for a mobile robot and investigate the validity and usefulness with a real robot. Our mobile robot with a fish-eye upward directed camera recognizes human pointing and navigates itself to the place the human points. A human user can navigate the mobile robot intuitively by simply pointing at desired position on a field with his hand. Experimental results show that the proposed method reasonably works and the user can control and navigate the robot intuitively. Adaptive control gain works well so that it reduces the different behavior of the robot according to distance from the robot and the pointing user and increases the operability and favorite of the user.

The current method forces a potential user to wear bright and uniform color jacket and glove to detect them by simple color extraction processing on vision system. It is not convenient for him/her to wear them every time to navigate the robot. A method extracting user region on the upward directed fish-eye camera image should be developed.



(a) adaptive gain



(b) fixed gain

Fig. 9. Errors with Adaptive/Fixed Gain Control during Navigation to a Fixed Point

At this moment, a potential user has to keep pointing at desired position until a robot reach it. As future work, the proposed method is extended so that a robot estimates direction and distance of the pointed place from a captured image of the pointing user. Nagai et al.[11] proposed a model of “joint attention” that enables a robot to look at an object which its caregiver watches. Their model trains a map from a caregiver’s face camera image to an input to head direction control. A similar idea might be applicable for our purpose, that is, a mobile robot learns a map from a captured image of a pointing user to a direction and distance of the pointing place as well.

REFERENCES

- [1] K. Y. Y. M. Fuminori Hibino, Yasutake Takahashi, “Human pointing navigation of mobile robot with spherical vision system,” in *Proceedings of Joint 5th International Conference on Soft Computing and Intelligent Systems and 11th International Symposium on Advanced Intelligent Systems*, vol. CD-ROM, 2010, pp. 217–221.
- [2] “Aibo official site,” <http://www.sony.jp/products/Consumer/aibo/>.
- [3] “Aldebaran robotics, the creators of nao — aldebaran robotics,” <http://www.aldebaran-robotics.com/>.
- [4] “irobot corporation: Home page,” <http://www.irobot.com/>.
- [5] “Ccp co.,limited — so-zi robotic vacuum cleaner [in japanese],” <http://www.ccp-jp.com/life/so-zi/>.



(a) Pointing

(b) Joystick



(c) Game-pad

Fig. 10. User Interfaces

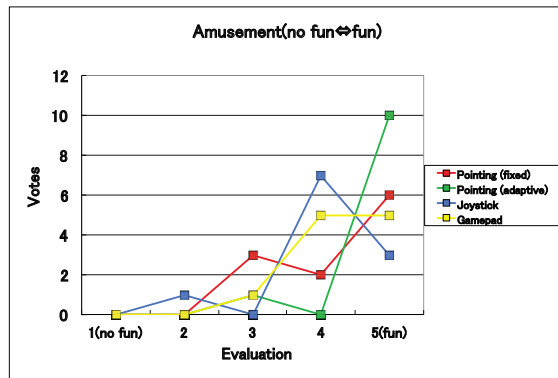


Fig. 11. Questionnaire Result: Amusement

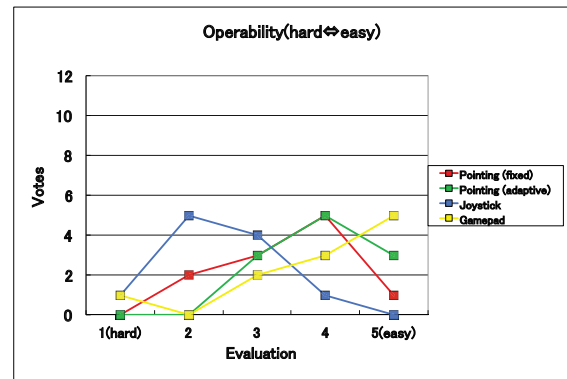


Fig. 12. Questionnaire Result: Operability

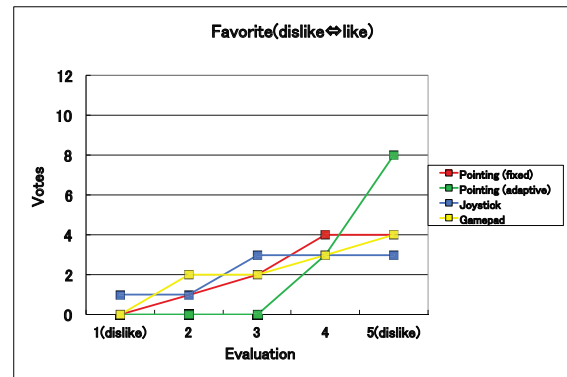


Fig. 13. Questionnaire Result: Favorite

technique,” in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Aug 2005, pp. 3195–3200.

- [11] Y. Nagai, M. Asada, and K. Hosoda, “Learning for joint attention helped by functional development,” *Advanced Robotics*, vol. 20, no. 10, pp. 1165–1181, 2006.

- [6] T. Nishimura, H. Yabe, and R. Oka, “A method of model improvement for spotting recognition of gestures using an image sequence,” *New Generation Computing*, vol. 18, no. 2, pp. 89–101, 2000.
- [7] Y. Iwai, H. Shimizu, and M. Yachida, “Real-time context-based gesture recognition using hmm and automaton,” vol. 0. Los Alamitos, CA, USA: IEEE Computer Society, 1999, p. 127.
- [8] N. Yoshiike and Y. Takefuji, “Object segmentation using maximum neural networks for the gesture recognition system,” *Neurocomputing*, vol. 51, pp. 213–224, apr 2003.
- [9] T. Fukuda, S. Ito, F. Arai, Y. Yokoyama, Y. Abe, K. Tanaka, and Y. Tanaka, “Navigation system based on ceiling landmark recognition for autonomous mobile robot-landmark detection based on fuzzy template matching (ftm),” *Intelligent Robots and Systems, IEEE/RSJ International Conference on*, vol. 2, p. 2150, 1995.
- [10] W. Jeong and K. M. Lee, “Cv-slam: a new ceiling vision-based slam